

CD

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :  C12N 15/10		A1	(11) International Publication Number: <b>WO 98/28416</b>  (43) International Publication Date: 2 July 1998 (02.07.98)									
<p>(21) International Application Number: PCT/DK97/00567</p> <p>(22) International Filing Date: 15 December 1997 (15.12.97)</p> <p>(30) Priority Data:</p> <table> <tr> <td>1471/96</td> <td>20 December 1996 (20.12.96)</td> <td>DK</td> </tr> <tr> <td>0592/97</td> <td>23 May 1997 (23.05.97)</td> <td>DK</td> </tr> <tr> <td>0935/97</td> <td>14 August 1997 (14.08.97)</td> <td>DK</td> </tr> </table> <p>(71) Applicant (for all designated States except US): NOVO NORDISK A/S [DK/DK]; Novo Allé, DK-2880 Bagsværd (DK).</p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants (for US only): BJØRNVAD, Mads, Eskelund [DK/DK]; Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd (DK). RASMUSSEN, Michael, Dolberg [DK/DK]; Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd (DK). JØRGENSEN, Per, Linå [DK/DK]; Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd (DK). BORCHERT, Torben, Vedel [DK/DK]; Novo Nordisk A/S, Novo Allé, DK-2880 Bagsværd (DK).</p> <p>(74) Common Representative: NOVO NORDISK A/S; Corporate Patents, Novo Allé, DK-2880 Bagsværd (DK).</p>		1471/96	20 December 1996 (20.12.96)	DK	0592/97	23 May 1997 (23.05.97)	DK	0935/97	14 August 1997 (14.08.97)	DK	<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> With international search report.</p>	
1471/96	20 December 1996 (20.12.96)	DK										
0592/97	23 May 1997 (23.05.97)	DK										
0935/97	14 August 1997 (14.08.97)	DK										
<p>(54) Title: IN VIVO RECOMBINATION</p> <p>(57) Abstract</p> <p>The present invention relates to a method for <i>in vivo</i> recombination of homologous DNA sequences situated on at least two different DNA structures. The method is a forced artificial evolution resulting in a DNA sequence encoding a polypeptide having an advantageous property.</p>												

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

<b>AL</b>	Albania	<b>ES</b>	Spain	<b>LS</b>	Lesotho	<b>SI</b>	Slovenia
<b>AM</b>	Armenia	<b>FI</b>	Finland	<b>LT</b>	Lithuania	<b>SK</b>	Slovakia
<b>AT</b>	Austria	<b>FR</b>	France	<b>LU</b>	Luxembourg	<b>SN</b>	Senegal
<b>AU</b>	Australia	<b>GA</b>	Gabon	<b>LV</b>	Latvia	<b>SZ</b>	Swaziland
<b>AZ</b>	Azerbaijan	<b>GB</b>	United Kingdom	<b>MC</b>	Monaco	<b>TD</b>	Chad
<b>BA</b>	Bosnia and Herzegovina	<b>GE</b>	Georgia	<b>MD</b>	Republic of Moldova	<b>TG</b>	Togo
<b>BB</b>	Barbados	<b>GH</b>	Ghana	<b>MG</b>	Madagascar	<b>TJ</b>	Tajikistan
<b>BE</b>	Belgium	<b>GN</b>	Guinea	<b>MK</b>	The former Yugoslav Republic of Macedonia	<b>TM</b>	Turkmenistan
<b>BF</b>	Burkina Faso	<b>GR</b>	Greece	<b>ML</b>	Mali	<b>TR</b>	Turkey
<b>BG</b>	Bulgaria	<b>HU</b>	Hungary	<b>MN</b>	Mongolia	<b>TT</b>	Trinidad and Tobago
<b>BJ</b>	Benin	<b>IE</b>	Ireland	<b>MR</b>	Mauritania	<b>UA</b>	Ukraine
<b>BR</b>	Brazil	<b>IL</b>	Israel	<b>MW</b>	Malawi	<b>UG</b>	Uganda
<b>BY</b>	Belarus	<b>IS</b>	Iceland	<b>MX</b>	Mexico	<b>US</b>	United States of America
<b>CA</b>	Canada	<b>IT</b>	Italy	<b>NE</b>	Niger	<b>UZ</b>	Uzbekistan
<b>CF</b>	Central African Republic	<b>JP</b>	Japan	<b>NL</b>	Netherlands	<b>VN</b>	Viet Nam
<b>CG</b>	Congo	<b>KE</b>	Kenya	<b>NO</b>	Norway	<b>YU</b>	Yugoslavia
<b>CH</b>	Switzerland	<b>KG</b>	Kyrgyzstan	<b>NZ</b>	New Zealand	<b>ZW</b>	Zimbabwe
<b>CI</b>	Côte d'Ivoire	<b>KP</b>	Democratic People's Republic of Korea	<b>PL</b>	Poland		
<b>CM</b>	Cameroon	<b>KR</b>	Republic of Korea	<b>PT</b>	Portugal		
<b>CN</b>	China	<b>KZ</b>	Kazakhstan	<b>RO</b>	Romania		
<b>CU</b>	Cuba	<b>LC</b>	Saint Lucia	<b>RU</b>	Russian Federation		
<b>CZ</b>	Czech Republic	<b>LJ</b>	Liechtenstein	<b>SD</b>	Sudan		
<b>DE</b>	Germany	<b>LK</b>	Sri Lanka	<b>SE</b>	Sweden		
<b>DK</b>	Denmark	<b>LR</b>	Liberia	<b>SG</b>	Singapore		
<b>EE</b>	Estonia						

**Title: In Vivo Recombination****Field of the invention**

The present invention relates to a method for *in vivo* recombination of homologous DNA sequences. The method is a forced artificial evolution resulting in a DNA sequence encoding a polypeptide having an advantageous property.

**Background of the invention**

Homologous recombination between DNA sequences placed on the same plasmid is well known to the skilled artisan (Weber and Weissmann, *Nucl. Acid. Res.* (1983) 11, 5661-5669). EP 252666 B1 (Novo Nordisk A/S), WO95/22625 A1 (Affymax Technologies N.V.) and WO9101087 discloses *in vivo* recombination of genes placed on the same plasmid.

EP 449923 B1 (Setratech) discloses a process of intergenic recombination *in vivo* of partially homologous DNA sequences. The recombination takes place in cells of which the enzymatic mismatch repair system is defective.

J. Biotechnol. (1991) 19, 221-240 discloses a process for inactivating a gene in the genome of *B. Amyloliquefaciens*. The process involves integration and excision of a pE194 mutant.

In Chemical Abstracts 118:161925 (1993) the minimal sequence length of homology necessary for recombination is determined. The method of determination involves integration of a pE194 derivative carrying a gluconase gene.

In J. Bacteriol. (1982) 152, 524-526 a plasmid pBD9, which comprises the two plasmids from *Staphylococcus aureus*, pE194 and pUB110, is disclosed. Other plasmids based on pE194 or pUB110 are disclosed in Molecular and General Genetics (1988), 213, 465-70, Molecular and General Genetics (1984), 195, 374-7, Plasmid (1981), 6, 67-77 and Genetika, (1986) 22, 2750-7.

In Yichuan Xuebao (1993), 20, 272-8 (see Chemical Abstracts 120:1670 (1994)), Chen et. Al. disclose the plasmid pNW102, a pE194 derivative, in which the thermostable  $\alpha$ -amylase gene from *Bacillus licheniformis* is inserted. pNW102 was transformed into the *Bacillus subtilis* strain BF 7658, followed by incubation at non-permissive temperature, allowing homologous

recombination between the  $\alpha$ -amylase genes of pNW102 and BF 7658. The  $\alpha$ -amylase produced of the recombinant strains shows the same characteristics as the  $\alpha$ -amylase from *Bacillus licheniformis*.

However, in the related art there is no indication of 5 that DNA sequences can be recombined *in vivo* by repeating integration and cross-out in order to obtain a gene encoding a polypeptide with improved characteristics, such as increased stability, improved specificity or higher activity.

Accordingly, a problem of the invention is to provide an 10 improved *in vivo* process of generating new DNA sequences.

#### Brief Description of the invention

The present invention is directed to a *in vivo* process of generating new DNA sequences. It has turned out that new DNA 15 sequences with improved properties can be generated in a fast and efficient way from at least two homologous DNA sequences by a process involving *in vivo* exchange of DNA between at least two vectors, each containing a homologous DNA sequence and an origin of replication.

20 The process of the invention is characterized by, that said exchange of said homologous sequences is done by repeating *in vivo* at least once, said integration and excision of one vector into the other.

Accordingly, in a first aspect the invention relates to 25 a process for *in vivo* recombination of homologous DNA sequences, comprising the following steps

- (a) incubating a cell containing at least two DNA structures, each comprising a DNA sequence and an origin of replication, under conditions which favour integration 30 of one of the DNA-structures into one of the others, thereby forming a hybrid DNA structure;
- (b) incubating the cell under conditions which favour crossing out from said hybrid DNA structure, thereby forming novel DNA structures, each comprising a recombined DNA 35 sequence and an origin of replication;
- (c) repeating steps a)-b) at least once.

One of the advantages of the *in vivo* recombination process of the invention is by repeating the exchange *in vivo* between the DNA sequences it is possible to obtain numerous different recombination patterns between said DNA sequences.

This is illustrated in figure 1 herein, which shows the result of *in vivo* recombination between the two DNA sequences Savinase and Savisyn performed as described in example 1.

In figure 1 Savinase/Savisyn recombined clones such as 10 clone no. 3, 8 12 are derived from more than one *in vivo* recombination event between the two DNA sequences.

Brief description of the drawing

Figure 1 illustrates the recombination of two DNA sequences, placed on the plasmids pTR, temperature resistant since 15 origin of replication is functional at 50°C, and pTS, temperature sensitive since origin of replication not functional over 45°C. Under non-permissive conditions (i.e. at 50°C), pTS integrates into pTR, forming a hybrid plasmid.

Figure 2 illustrates the cross out from the hybrid plasmid, forming two new DNA sequences, each placed on a plasmid. Repeated crossing in and out may be done by e.g. temperature cycling or simply by keeping temperature pressure and selecting 20 for resistance of pTS (i.e.  $\text{Erm}^R$ ).

Figure 3 illustrates the plasmids which are obtained after 25 repeated integration and crossing out events.

Further, figures 1 and 2 illustrate a DNA structure system suitable for *in vivo* measurement of hybrid structure formation, and suitable for performing a forced cross out of the DNA 30 hybrid structure.

Fig. 1: DNA structure pTS:

Suitable active promoter directing expression of active protein

$\text{Erm}^R$ : antibiotic resistance gene

DNA structure pTR comprises:

Not any active promoter to direct transcription of GFP

## GFP: Green Fluorescent Protein

The  $\text{Erm}^R$  antibiotic resistance gene in pTS is active due to the active promoter.

5 The GFP gene in pTR is inactive (i.e. the gene is non-transcribed) due to there is not an active promoter to drive expression of the gene.

Fig. 2: After formation of the hybrid structure according to the invention GFP is now active (driven by the promoter). This makes it possible to *in vivo* measure the hybrid structure 10 formation (see below for further details).

In contrary the  $\text{Erm}^R$  antibiotic resistance gene is now inactive (i.e. the gene is non-transcribed) due to there is not an active promoter to drive expression of the gene. Forced cross out of the DNA hybrid structure is performed by incubating the 15 cells under  $\text{Erm}^R$  antibiotic pressure, thereby forcing a cross out of the DNA hybrid structure, since the cell is only able to express the  $\text{Erm}^R$  antibiotic resistance gene after the crossing out of the DNA structure has taken place.

Figures 4 and 5: Illustrate the two vectors pSX120 (fig. 20 4) and pMB430 (fig. 5) used as described in example 1 to *in vivo* recombine the two DNA sequences Savinase and Savisyn.

pSX120 comprises i) an open reading frame encoding Savinase, ii) an active promoter directing transcription of Savinase, and iii) a gene conferring resistance to chloramphenicol 25 ( $\text{Cam}^R$ ).

pMB430 comprises i) a Savizyn DNA fragment, ii) a terminator (resulting in no transcription through the regions following this), iii) a gene conferring resistance to erythromycin ( $\text{Erm}^R$ ), and iv) an open reading frame encoding Green Fluorescent 30 Protein (GFP).

Fig. 6: Illustrates the results of Savinase/Savisyn *in vivo* recombination performed as described in example 1.

The schematic representation shows 12 of the shuffled/recombined clones.

35 Horizontally are given the positions of the restriction sites unique for Savizyn.

The letter "S" denotes that the sequence at this position is identical to the Savinase sequence and the letter "Z" denotes

that the sequence stems from Savizyn. All of the sequenced clones are different in their pattern of recombination.

Definitions

5 As used herein, the following terms have the following meanings:

The term "DNA sequence" includes any DNA sequence which is desirable to modify according to the invention e.g. a DNA sequence encoding, a polypeptide, for example an enzyme, pharmaceutically active polypeptides, e.g. insulin growth hormone human hormones or growth regulators, and sequences such as promoters, transcription or translation regulators and other intracellular polynucleotides and polypeptides.

The term "homologous" means the DNA sequences have some 15 stretches of identical nucleotides, allowing recombination *in vivo*. Said stretches within the DNA sequences are at least 15 base pair long, more preferably at least 25 base pair long, even more preferably at least 50 base long, and most preferably at least 150 base pair or more.

20 The term "homologous" comprises stretches of DNA sequences that are from 60% identical to DNA sequences that differs in only one nucleotide. It is preferred that the stretches of DNA sequences are more than 70% identical, more preferably at least 80%, especially more than 90% identical, and 25 more preferred that the stretches of DNA sequences are 95% identical, most preferred at least 97%.

Within the stretches of the DNA sequences of interest, the DNA sequence homology referred to above is determined as the degree of identity between two sequences indicating a derivation 30 of the first sequence from the second. The homology may suitably be determined by means of computer programs known in the art, such as GAP provided in the GCG program package (Program Manual for the Wisconsin Package, Version 8, August 1994, Genetics Computer Group, 575 Science Drive, Madison, Wisconsin, USA 35 53711) (Needleman, S.B. and Wunsch, C.D., (1970), Journal of Molecular Biology, 48, 443-453). Using GAP with the following settings for DNA sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the homologous DNA sequences

of interest exhibits a degree of identity preferably as described above.

The term "Recombined DNA sequence" means the DNA sequence which is a result of a recombination event between at least two DNA sequences. The recombination event may be a single recombination event or may be first a recombination event between at least two DNA sequences followed by a cross out at a different site than the first recombination event.

The term "DNA structure" means a DNA molecule, e.g. a vector, a plasmid, a chromosome (e.g. a *Bacillus subtilis* chromosome), a virus, a phage, a transposon or a genome.

The term "vector" means any linear or circular DNA molecule that is able to contain a DNA sequence to be recombined and an origin of replication. The term includes any vector known to the skilled artisan e.g. a plasmid or a phage. The DNA sequence is optionally-but not necessarily- under control of a promoter.

The term "a cell" used herein in connection with a *in vivo* recombination process of the invention is intended to cover both "a cell" and "a cell population".

The term "transferring" means that the DNA structure is introduced into a cell. The DNA structure may be introduced using techniques known in the art for direct introduction of DNA, e.g. by use of electroporation, transformation of competent cells, protoplast transformation, transfection, transduction, conjugation or ballistic transformation.

The term "incubating" (or cultivating, growing) means that the cell is treated in such a way that the normal cell functions are active, especially the functions necessary for recombination. The cell can be incubated on a solid substrate (agar petri disc) or the cell can be incubated under submerse conditions. It is preferred that the cell is incubated in a test tube, e.g. an eppendorph test tube. In the case of a temperature-sensitive origin of replication, the test tube is conveniently placed in an automatic termocycler. Further the cell may be incubated in standard shake-flaks.

The term "origin of replication" (ori) means a region where replication of a DNA structure is initiated. The term

functional in connection with origin of replication means that replication can be initiated from the origin. The term non-functional means that no replication is initiated under the given conditions, or that the initiation of replication is reduced, e.g. to a level insufficient for survival of a cell containing a DNA structure with a marker gene and the ori, under appropriate selection pressure.

The term "temperature sensitive" in connection with origin of replication or vector means the vector is unable to replicate at e.g. increased temperatures, i.e. non-permissive conditions, which yet permit growth of the parent cell.

The term "temperature resistant" in connection with origin of replication or vector means the vector is able to replicate at increased temperatures.

15 The term "DNA library" is a library comprising numerous different DNA sequences. A DNA library may be made according to standard procedures known in the art, such as e.g. by in vitro DNA shuffling/DNA recombination (WO 95/17413, WO 95/22625) and/or Error-prone PCR and cassette mutagenesis.

20 A DNA library herein may comprise as low as 2-20 different sequences, as e.g. a library where only one amino acid is varied. Further a DNA library may comprise numerous different DNA sequences, e.g. up to  $10^{10}$  different sequences.

25 Detailed Description of the Invention

The present invention comprises a process for *in vivo* recombination of homologous DNA sequences, comprising the steps

- (a) incubating a cell containing at least two DNA structures, each comprising a DNA sequence and an origin of replication, under conditions where i) at least one origin of replication is non-functional and ii) the cell is under a selective pressure which only allows growth of the cell if it contains the DNA structure which is having the non-functional origin (i.e. favouring an integration of one of the DNA-structures into one of the others, thereby forming a hybrid DNA structure);
- (b) changing of said conditions to conditions, where fewer origins of replication are non-functional than in step

- a) (i.e. favouring a crossing out from said hybrid DNA structure);
- (c) repeating steps a)-b) at least once.

5 The recombined DNA sequence will encode new polypeptides having new properties. It is then possible to incorporate a selection or screening system in the process of the invention in order to select and/or screen for desired properties encoded by of or more of the recombined DNA sequence(s).

10 The number of repetition(s) in step (c) above may be one repetition or it may be more such as 5-10 repetitions or 50-100 or even higher numbers of repetitions.

15 It is preferred that integration is forced (or made favorable for survival of the DNA structure) by incubation of the cell under such conditions where the origin of replication of a DNA structure is non-functional, and that cross out is forced by incubation of the cell under such conditions where the origin of replication of a DNA structure again is functional.

20 In another preferred embodiment the routine depicted above in the steps (a) to (c) it is preferred that the cells are kept under constant conditions (e.g. constant temperature and selection pressure) where spontaneous cross-in and cross-out happens. That is the shuffling procedure in the cells is repeated without changing the conditions repeatedly.

25 In a number of DNA structures, such as in many *Bacillus* DNA structures (vectors) (*Bacillus subtilis* and Other Gram-Positive Bacteria, Sonensheim et al., 1993, American Society for Microbiology, Washington D.C.) the origin of replication comprise a ori(+) and a ori (-). ori(+) controls where replication 30 of the first DNA strand starts, and ori(-) controls the initiation point for replication on the second strand (*Bacillus subtilis* and Other Gram-Positive Bacteria, Sonensheim et al., 1993, American Society for Microbiology, Washington D.C.).

35 In an embodiment of the process of the invention, at least one DNA structure does not comprise ori(-) (i.e. only comprise ori(+)). Preferably the DNA structure which does not comprise ori(-) is a DNA structure which origin (i.e. herein

ori(+)) is functional during the hybrid DNA structure formation performed according to the invention.

One advantage of this embodiment of the invention is when only ori(+) is functional in said DNA structure, this DNA 5 structure is in a single-stranded state for relatively longer period, compared to a DNA structure comprising both ori(+) and ori(-). This prolonged single-stranded state facilitate the hybrid DNA structure formation according to a process of the invention.

10 Preferably, the DNA structure which does not comprise ori(-) as described above is a *Bacillus* structure, more preferred a *Bacillus* vector.

15 In order to illustrate this concept, an example is a modified pTR (TR: temperature resistance) *Bacillus* vector which does not comprise ori(-) (see figure 1). If the hybrid formation is performed at 50°C only the origin of pTR is functional (origin of pTS (TS: temperature sensitive) is not functional) (see fig. 1,2). The DNA structure pTR only comprise ori(+) and is therefore in a single-stranded state for relatively longer 20 period, which facilitate the recombination event and consequently hybrid formation with the pTS DNA structure (see fig. 1,2).

25 It is more preferred that one DNA structure is a vector which is able to replicate under certain (permissive) conditions and unable to replicate under other (non-permissive) conditions. The vector may, for instance, be one which is temperature-sensitive for replication. Thus, the vector may be one which is unable to replicate at increased temperatures, which yet permit growth of the parent cell. The cell is initially incubated at a 30 temperature permitting vector replication and subsequently after integration into the other DNA structure may have taken place, incubated at a temperature which does not permit vector replication so that the vector is lost from the cells unless integrated.

35 The vector may further comprise a selectable marker. In this case, the incubation at the non-permissive temperature may be conducted under selective conditions to ensure that only

cells containing the integrated vector (which includes the DNA sequence and the selectable marker), will survive.

Further, it is preferred to construct a system where it is possible *in vivo* to measure formation of the hybrid structure.

Preferably, this is done by constructing a system wherein an *in vivo* screenable or selectable protein is expressed only after the hybrid structure has been formed.

Preferably, said screenable protein is a protein which is fluorescent under suitable exposure, and in particular said fluorescent protein is Green Fluorescent protein (GFP) or variants thereof.

Green Fluorescent protein (GFP) or variants thereof and how the GFP's are activated to fluoresce are described in the art and reference is made to (Crameri et al. *Nature Biotechnology* 14:315-319 (1996); and Cormack et al. *GENE* 173:33-38 (1996)) for further details.

One of the advantages of using such a screenable protein which is fluorescent under suitable exposure of light, and in particular GFP or variants thereof, is that it is then possible by e.g. fluorescent-activated cell sorting (FACS) technology to selectively sort out cells which comprise said hybrid structure. This may eliminate background of cells, which do not comprise said hybrid structure after performing of step a) according to the first aspect of the invention.

FACS is a widely known screening technology and reference is made to (Cormack et al. "FACS-optimized mutants of the green fluorescent protein (GFP)" *GENE* 173:33-38 (1996)) for further details.

In order to illustrate a suitable system enabling *in vivo* measurement of the formation of the hybrid structure, an example of a suitable system is shown in figure 1 and 2.

Alternatively to the use of a screenable protein as discussed above, it may be advantageously to use a selectable marker, wherein said selectable marker is only transcriptionally active when the hybrid structure is formed.

Further, it is preferred to that the crossing out from the hybrid DNA structure, in step b) of the first aspect of the invention, is a forced crossing out event.

Forced crossing out from the hybrid structure may be 5 performed by:

- i) constructing a system wherein the DNA hybrid structure does not express a selectable marker, and at least one of the at least two DNA structures in step a) of first aspect is able to express said selectable marker;
- 10 ii) incubating the cell under selectable conditions, thereby forcing a cross out of the DNA hybrid structure, since the cell is only able to express the marker gene after the crossing out of the DNA hybrid structure has taken place.

15 Preferably, selectable marker gene mentioned above is expressed on a DNA structure where the origin of replication is non-functional during a forced integration of the DNA structures into a DNA hybrid structure performed as described above, and preferably the selectable marker gene is an antibiotic resistance gene. See below for further description of suitable selectable marker genes.

In order to illustrate the concept of forced crossing out from the hybrid DNA structure, an example of a suitable system is shown in figure 2 1 and 2.

25 The selectable marker may be any marker known in the art, for instance a gene coding for a product which confers antibiotic resistance to the cell, which confers prototrophy to an auxotrophic strain, or which complements a defect of the host, (e.g. dal genes introduced in a dal- strain; cf. B. Diderichsen 30 (1986), *Bacillus: Molecular Genetics and Biotechnology Applications*, A.T. Ganesan and J.A. Hoch, Eds., Academic Press, pp. 35-46. The selectable marker may, e.g., be excised from a known source or present on a vector, e.g. a plasmid, used for the construction of the DNA structure to be used in the method of the 35 invention.

This selection may be accomplished by growing the cells under selection pressure for the selection marker encoded by the marker gene.

The DNA structure (in which the vector is to be integrated) may contain a gene encoding any selectable marker, e.g. of any type as described above in connection with the marker optionally carried by the vector to be used in the method 5 of the invention. Thus, the marker gene may encode an antibiotic resistance such as resistance to kanamycin, tetracyclin, ampicillin, erythromycin, chloramphenicol, or a resistance to various heavy metals such as selenate, antimony or arsenate.

In a preferred embodiment, the cell used in the process 10 is defective or transitorily inactivated in the enzymatic mismatch repair system or has a reduced level of mismatch repair. In the synthesis of DNA, errors can occur and the resulting non-complementary pairs of bases are called mismatches. A process for the correction of errors (mismatches) in DNA exists. In *E. coli* the errors are very rapidly and accurately detected by two 15 enzymes (MutS and MutL) enabling a third enzyme (MutU) to unwind the two DNA strands and a fourth enzyme (MutH) to cut the neo-synthesized strand on a sequence of DNA (GATC) itself methylated later. For a closer explanation, reference is made to EP 449923 20 (Setratech), the contents of which hereby are incorporated by reference.

In another preferred embodiment the cell used in the process is impaired or hampered in the major recombination enzymes such as AddAB. It has been shown that a strain impaired in 25 the AddAB genes shows a higher frequency of homologous recombination. Both legitimate and illegitimate recombination frequency are increased. But especially the recombination frequency for illegitimate recombination is favoured when working with a strain impaired or hampered in the AddAB ATP-dependent nuclease 30 (Meima, R ; Hajema, BJ; Dijkstra, H; Haan, GJ; Venema, G; Bron, S (1997) Role of enzymes of homologous recombination in illegitimate plasmid recombination in *Bacillus-subtilis*. JOURNAL OF BACTERIOLOGY Vol. 179 , No. 4 pp. 1219-1229.)

In another preferred embodiment the cell used in the 35 process is over-expressing one of the key recombination enzymes, recA. The recA enzyme is involved in homologous recombination by promoting renaturation of ssDNA to form the heteroduplex molecules necessary for the recombination event and initiates the exchange

of DNA strands. (PROGRESS IN NUCLEIC ACID RESEARCH AND MOLECULAR BIOLOGY Vol. 56 , pp. 129-223 (1997)). The recA over-expression can be achieved by either classical mutagenesis of the host strain or cloning of a highly expressed recA gene.

5 In another preferred embodiment the cell used in the process is subjected to UV, X-ray or DNA damaging agents to increase recombination frequency between the two homologous genes to be shuffled. Stress like UV, X-ray or DNA damaging agents is known to induce the recombination enzymes via the SOS response and 10 activate enzymes involved in recombination processes like recA. (BACILLUS SUBTILIS AND OTHER GRAM-POSITIVE BACTERIA, pp.529-537 (1993)).

In another preferred embodiment the pTR based plasmid used as a carrier of one of the genes to be shuffled is replaced by 15 either a low-copy version of pTR or another B. subtilis compatible low-copy plasmid like pWVO1, (MOLECULAR AND GENERAL GENETICS Vol. 249 , No. 1 pp. 43-50 (1995 )), pTA1060, (JOURNAL OF BACTERIOLOGY Vol. 142 , No. 1 pp. 315-8 (1980)) or pAMb1, (JOURNAL OF BACTERIOLOGY Vol. 157 , No. 2 pp. 445-53 (1984)).

20 In a further preferred embodiment the one of DNA construct comprising one the genes to be shuffled is situated on the chromosome of the cell and another DNA construct, comprising another gene of interest to be recombined is a plasmid. When performing an in vivo recombination process of the invention the 25 plasmid will be recombined with the DNA construct on the chromosome. Said plasmid may e.g. be a temperature sensitive plasmid.

The advantages of using a low-copy pTR, or a system where 30 one of the DNA constructs is situated on the chromosome as described above, is that the hybrid DNA construct after recombination of the two homologous genes will be the predominant form of the genes in the cell. A high-copy background of non-recombined pTR plasmids will inevitably reverse the gene hybrids formed in the first round back to wildtype after the second round 35 of recombination and therefore prohibit any shuffling of the two genes. When using the chromosome as a carrier of one of the genes no wildtype background can coexist when the pTS has integrated via the homologous genes.

It is preferred that the cell used in the process is a cell suitable for large scale production of the product of the recombined DNA sequence. It is preferred that the cell is a bacterial cell, more preferred a *Bacillus* cell, most preferred a 5 *Bacillus subtilis* cell.

The DNA sequences to be recombined may be isolated/cloned from organisms containing the sequences in a manner known per se. For instance, a suitable oligonucleotide probe may be prepared on the basis of a first DNA sequence of interest in order to 10 obtain a second DNA sequence homologous to the first sequence. Alternatively, the DNA sequence can be amplified by PCR based on primers designed according to knowledge of conserved regions.

I different embodiments of the present content is used the term "a set of homologous DNA sequences ("set A" "set B", 15 "setC" etc.)", e.g. in relation to an embodiment of the invention relating to a process according to the invention, wherein the homologous DNA sequences are comprised of two or more different set of homologous DNA sequences ("set A" "set B", "setC" etc.), and wherein each of said set of homologous 20 sequences ("set A" "set B", "setC" etc.) are inserted in a different DNA construct ("DNA construct A", "DNA construct B", "DNA construct C" etc.) according to the invention.

This term is used to illustrate that an *in vivo* recombination process of the present invention may be used to 25 recombine numerous "sets" of homologous DNA sequences. Each individual "set" of homologous DNA sequence(s) may individually comprise from only one to numerous (e.g. a "set" comprising a DNA library) individual homologous DNA sequence(s).

Each of said "sets" of homologous DNA sequences ("set A" 30 "set B", "setC" etc.) may then be inserted in a different DNA construct ("DNA construct A", "DNA construct B", "DNA construct C" etc.) according to the invention. Each of said different DNA construct are preferably different on important parameters, according to the invention, such as different origin of 35 replication, different selectable markers etc.

All of said DNA constructs ("DNA construct A", "DNA construct B", "DNA construct C" etc...) may then be transferred into a cell, according to the invention, and when performing a

process for in vivo recombination of the invention, then each of the individual DNA constructs ("DNA construct A", "DNA construct B", "DNA construct C" etc..) may then recombine randomly according to the invention. After a few repetitions of a 5 process, of the invention, this will give a DNA population, in the cell population, comprising numerous different combinations of different recombined DNA sequences, recombined randomly between all "sets" of homologous DNA sequences used in the process.

10 Further, the term "a set of homologous DNA sequences ("set A" "set B", "setC" etc.)" is used herein to illustrate one of the advantages of an in vivo recombination process of the invention, which is the flexibility of the system to recombine both set of homologous DNA sequences comprising one, few and/or 15 numerous (e.g. big DNA libraries) of individual homologous DNA sequences.

Accordingly, an embodiment of the invention relates to process according to the invention, wherein said sets of homologous DNA sequences each only contain one DNA sequence 20 (i.e. recombination of two or more homologous sequences).

Further in embodiments of the invention the homologous DNA sequences to be recombined are derived from a pre-made DNA library and/or derived from individually different pre-made libraries.

25 Homologous DNA derived from such DNA libraries are inserted into a vector according to the invention (i.e. DNA re-presenting the DNA library is ligated (inserted) into a vector). As described herein the process of the invention comprise at least two DNA structures (e.g. vectors), which are different in 30 a way which allow forced in vivo recombination of said vectors.

Homologous DNA representing DNA libraries may be inserted into only one and/or both of said different vectors.

I.e. if for instance homologous DNA representing one DNA library is inserted (ligated) into one of said different vectors 35 and homologous DNA representing another DNA library is inserted into the other vector this will allow fast and efficient recombination between two different libraries by an in vivo recombination process according to the invention. To illustrate

this concept, DNA representing one library may be inserted into a pTR vector (see figure 1) and DNA representing another library may be inserted into a pTS vector (see figure 1). Just a few repeating in vivo recombination events performed by a process according to the invention will then give a cell population representing a new library which is a combination of the two libraries mentioned above.

Accordingly, embodiments of the invention relates to a process according to the invention, wherein at least one of the sets of homologous DNA sequences is derived from a pre-made DNA library, and at least one of said sets of homologous DNA sequences is a DNA sequence which is homologous to said DNA library (i.e. resulting in a process for in vivo recombination of said DNA library to a DNA sequence which are homologous to said DNA library); or

a process according to the invention, wherein the sets of homologous DNA sequences is derived from the same pre-made DNA library (i.e. resulting in a process for in vivo recombination of said DNA library); or

a process according to the invention, wherein the sets of homologous DNA sequences are derived from at least two different pre-made DNA libraries (i.e. resulting in a process for in vivo recombination of at least two different DNA libraries).

In order to obtain even further recombination between the DNA sequences it may be preferred to perform an isolation of the DNA structures after a first round of in vivo recombination of said DNA sequences performed according to the invention. Said isolated structures are then re-transformed into the cells and a further process round of in vivo recombination is then performed.

The advantages of this step is illustrated by following example:

Set A is a vector-A wherein DNA sequences a, b are introduced individually in said vector A (i.e. the vector A population comprises a vector with DNA sequence a, and a vector with sequence b); and

Set B is a vector B wherein DNA sequences c, d are introduced individually in said vector B (i.e. the vector B population comprises a vector with DNA sequence c, and a vector with sequence d).

5 Vector A and B are then transformed into a cell population and *in vivo* recombined according to the invention. This will give recombination between DNA sequences ac, ad, bc, bd. The vectors A, and B comprising said recombined sequences is isolated, and re-transformed into said cell population. *In vivo* 10 recombination according to the invention will then give all possible recombination of said four (a,b,c,d) DNA sequences, such as recombination of abc, abdc, etc.

In even further embodiments the invention relates to a process according to the invention, wherein two sets of homologous DNA sequences ("set A" and "set B") are recombined *in vivo*, comprising the steps

- a) inserting sequence(s) from "set A" in a vector containing a temperature resistant origin of replication and a gene for a first marker;
- 20 b) inserting sequence(s) from "set B" in a vector containing a temperature sensitive origin of replication and a gene for a second marker;
- c) transferring the vectors into a cell;
- d) incubating the cell at a temperature where both origins 25 of replication are functional, optionally under a selection pressure that favours cells with both marker genes;
- e) shifting the temperature to a temperature, where only the temperature resistant origin of replication is functional;
- 30 f) repeating steps d)-e) at least once;

or

a process according to the invention, wherein two sets 35 of homologous DNA sequences ("set A" and "set B") are recombined *in vivo*, comprising the steps

- a) inserting sequence(s) from "set A" in a vector containing a temperature resistant origin of replication and a gene for a first marker;
- 5 b) inserting sequence(s) from "set B" in a vector containing a temperature sensitive origin of replication and a gene for a second marker;
- c) transferring the vectors into a cell;
- 10 d) incubating the cell at a temperature where only the temperature resistant origin of replication is functional, optionally under a selection pressure that favours cells with both marker genes;
- e) optionally, shifting the temperature to a temperature, where both origins of replication are functional;
- 15 f) repeating steps d)-e) at least once.

15 In a further aspect the invention relates to a method for producing one or more recombinant protein(s) having a desired biological activity comprising:

- 20 a) performing a process for in vivo recombination of homologous DNA sequences (the process is performed according to the invention as described herein), wherein at least one of the DNA structure(s) is an expression vector;
- b) expressing the numerous different recombinant polypeptides encoded by the numerous different recombined DNA sequences from step a); and
- 25 c) screen or select the numerous different recombinant proteins from step b) in a suitable screening or selection system for one or more recombinant polypeptides(s) having a desired activity.

30 Preferably, the expression vector in step a) above is a DNA structure which is unable to replicate (e.g. having a non-functional origin of replication) during a forced hybrid DNA structure formation (i.e. integration of said structures into each other) performed according to a process for in vivo recombination of homologous DNA sequences according to the invention. This will improve the probability that the expression

vector, after a crossing-out event of said hybrid-structure according to the invention, comprise recombined homologous DNA, since the only way said expression vector may have survived the process of forced hybrid formation is by actually recombine with 5 another DNA structure, which is able to replicate under the conditions used to force said hybrid-structure formation.

After the process for in vivo recombination have been performed (step a) above), and before the expression and screening steps b) and c) above, it may be advantageously to 10 make a cell population mainly comprising only the expression vector (i.e. other DNA construct(s) are partly or nearly completely removed from the cell population). This is particular advantageously when the expression vector is unable to replicate during the forced hybrid DNA structure formation as described 15 immediately above, since such a cell population will have a very limited background of non-recombined DNA sequences of interest, which is advantageously in the subsequent screening step (step c) above).

Making a cell population mainly comprising only the 20 expression vector may be performed according to standard procedures known in the art, such as growing the cell population at conditions selectively favouring replication of the expression vector, as compared to the other DNA structures; or selectively purify said expression vector from the cells and subsequently 25 transfer this into a new cell population.

Expression the recombinant polypeptides encoded by the recombinant sequence in step a) above may be performed by use of standard expression vectors, and/or a standard expression vector may be modified in order to comprise elements specially suitable 30 for a in vivo recombination process according to the invention.

A suitable screening or selection system, to screen or select the numerous different recombinant proteins from step b) above, will depend on the desired biological activity.

A number of suitable screening or selection systems to 35 screen or select for a desired biological activity are described in the art. Examples are:

Strauberg et al. (Biotechnology 13: 669-673 (1995), which describe a screening system to screen for Subtilisin variants having a Calcium-independent stability;

5 Bryan et al. (Proteins 1:326-334 (1986)), which describe a screening assay to screen for Protease having a enhanced thermal stability; and

PCT-DK96/00322 which describe a screening assay to screen for lipases which are having an improved wash performance in washing detergents.

10 A preferred embodiment of the invention comprise screening or selection of recombinant protein(s), wherein the desired biological activity is improved performance in dish-washing or laundry detergents. Examples of suitable dish-wash or laundry detergent are disclosed in PCT/DK96/00322 and WO 95/30011.

15 The invention also comprises a set (at least two) of vectors, comprising at least two different origins of replication able to function in the same cell under permissive conditions, and at least two homologous DNA sequences encoding polypeptides. In a preferred embodiment, at least one origin of 20 replication is temperature sensitive. The vector may preferable be a plasmid.

The invention further comprises a vector system, comprising at least

- a) a vector A containing a DNA sequence encoding a polypeptide and an origin of replication, and
- 25 b) a vector B containing a DNA sequence homologous to the DNA sequence in vector A, and an origin of replication that differs from the origin of replication in vector A. The vectors may preferable be plasmids.

30 It is preferred that the DNA sequence encodes polypeptides with a biological activity, especially enzymes and in particular enzymes such as an amylase, lipase, cutinase, cellulase, oxidase, phytase, and a protease.

35 The invention further comprises a new, recombinant DNA sequence, characterised in that has be produced by a process according to the invention, or it is encoding a recombinant protein having a desired biological activity, wherein said recombi-

nant protein having a desired biological is produced by the method according to method for producing one or more recombinant protein(s) having a desired biological activity according to the invention.

5 The invention further comprises a process for producing a polypeptide, especially an enzyme, comprising i) growing a cell expressing a polypeptide encoded by a recombined DNA sequence according to claim 38, and ii) purifying said expressed polypeptide.

10 The invention further comprises a new polypeptide, especially an enzyme, characterised in that it is produced by a process according to the invention.

15 The invention further comprises a polypeptide, especially an enzyme, characterised in that it is encoded of a re-combined DNA sequence according to the invention.

20 Polypeptides of the present invention include peptides and proteins, and optionally glycosylated forms thereof. Examples are enzymes, pharmaceutically active polypeptides such as analogues of human or mammal polypeptides, e.g. insulin, growth hormone, human hormones or growth regulators.

The invention is described in further detail in the following examples which are not in any way intended to limit the scope of the invention.

## 25 EXAMPLES

### MATERIALS AND METHODS

#### strains.

30 *E.coli*: SJ2 (Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of *aldB*, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. *J. Bacteriol.*, 172, 4315-4321) Electrocompetent cells prepared and transformed using a Bio-Rad GenePulser™ as  
35 recommended by the manufacturer.

*B.subtilis* DN1885 Diderichsen, B., Wedsted, U., Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of *aldB*, which encodes

alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. J. Bacteriol., 172, 4315-4321). Competent cells were prepared and transformed as described by Yasbin, R.E., Wilson, G.A. and Young, F.E. (1975) Transformation and transfection in 5 sogenic strains of *Bacillus subtilis* : evidence for selective induction of prophage in competent cells. J. Bacteriol., 121:296-304.

*B. subtilis* PL1801. This strain is the *B. subtilis* DN1885 with disrupted apr and npr genes (Diderichsen, B., Wedsted, U., 10 Hedegaard, L., Jensen, B. R., Sjøholm, C. (1990) Cloning of aldB, which encodes alpha-acetolactate decarboxylase, an exoenzyme from *Bacillus brevis*. J. Bacteriol., 172, 4315-4321).

**Plasmids:**

15 pSX120: *B. subtilis* expression vector (Described in EP 405901, and WO 91/00345). This vector comprise the wild-type Savinase gene.  
pE194: (Horinouchi, S and Weisblum, B., 1982, J. Bacteriol. 150:804-814).  
20 pSX222: *B. subtilis* expression vector (Described in WO 96/34946, and WO 92/19729). This vector comprise the synthetic Savinase gene herein called Savisyn. In the Savisyn gene is introduced a large number of restriction sites as compared to the wild-type Savinase gene comprised in pSX120.

25

pUC19: Yanisch-Perron, C., Vieria, J. and Messing, J. (1985) Gene 33, 103-119.

30 pF64L-S65T-GFP: is the *E.coli* plasmid encoding the DNA fragment encoding the F64L-S65T-GFP described in WO 97/11094.

pMUTIN-4-MCS: Plasmid can be obtained from Laboratoire de Génétique Microbienne, Institut National de la Recherche Agronomique, 78352 Jouy en Josas - CEDEX, France.

35

Solutions/Media

TY and LB agar (as described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995).

LBPG is LB agar supplemented with 0.5% Glucose and 0.05 M s potassium phosphate, pH 7.0

LBPGSK plates are LBPG media with agar and 1% of skimmed milk.

Antibiotics are used in different medias in concentrations of 6 µg/ml Chloramphenicol, 10 µg/ml Kanamycine and 5 µg/ml og 10 Erythromycine.

TE-buffer (as described in Ausubel, F. M. et al. (eds.) "Current protocols in Molecular Biology". John Wiley and Sons, 1995).

#### 15 General molecular biology methods

DNA manipulations and transformations were performed using standard methods of molecular biology (Sambrook et al. (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor lab., Cold Spring Harbor, NY; Ausubel, F. M. et al. (eds.) "Current 20 protocols in Molecular Biology". John Wiley and Sons, 1995; Harwood, C. R., and Cutting, S. M. (eds.) "Molecular Biological Methods for Bacillus". John Wiley and Sons, 1990).

Enzymes for DNA manipulations were used according to the specifications of the suppliers.

25

#### Enzymes for DNA manipulations

Unless otherwise mentioned all enzymes for DNA manipulations, such as e.g. restriction endonucleases, ligases, polymerases, PCR polymerases etc., are obtained from New England 30 Biolabs, Inc.

#### Construction of the temperature sensitive shuffling plasmid

pMB430 was constructed in a five step cloning process as depicted in the following, the entire DNA sequence is found in 35 SEQ ID No. 1.

The temperature sensitive plasmid pE194 was modified by introducing multiple cloning sites around the unique *Cla*I site of the pE194 plasmid and thus destroying the *Cla*I site. This introduction was done by PCR using the primers #22899 and #24039 to amplify the whole plasmid and introducing several unique cloning sites.

The PCR was performed as follows:

1  $\mu$ l of pE194 Qiaquick plasmid prep was used for 50  $\mu$ l PCR amplification in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01 % (w/v) gelatin) containing 200  $\mu$ M of each dNTP, 2.5 units of AmpliTaq polymerase (Perkin-Elmer, Cetus, USA) and 100 pmol of each primer:

#22899

5'-GCAGCTGGATCCGGGCCGCGAATTCTCGCTTTGCGCAGTCGGC-3'

15

#24039

5'-GCAGCTGGATCCGGGCCGGGAAGCTTCGATTCACAAAAATAGGCACACG-3'

The PCR reactions were performed using a DNA thermal 20 cycler (Landgraf, Germany). One incubation at 94°C for 1 min followed by forty cycles of PCR performed using a cycle profile of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, and extension at 72°C for 2 min. Five- $\mu$ l aliquots of the amplification product was analyzed by electrophoresis in 0.7 % 25 agarose gels (NuSieve, FMC).

#### Subcloning of PCR fragment

Forty-five- $\mu$ l aliquots of the PCR products generated as described above were purified using QIAquick PCR purification 30 kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50  $\mu$ l of 10mM Tris-HCl, pH 8.5. Fifty - $\mu$ l of the purified PCR fragment was digested with BamHI, electrophoresed in 0.8 % low gelling temperature agarose (SeaPlaque GTG, FMC) gels, the relevant fragment was excised 35 from the gels, and purified using QIAquick Gel extraction Kit (Qiagen, USA) according to the manufacturer's instructions. The isolated DNA fragment was then ligated to BamHI digested pUC19 and the ligation mixture was used to transform *E. coli* SJ2.

Cells were plated on LB agar plates containing Erythromycin (200 µg/ml) supplemented with X-gal (SIGMA, USA) (5-Bromo-4-chloro-3-indolyl alpha-D-galactopyranoside, 50 µg/ml).

5 Identification and characterisation of positive clones

The transformed cells were plated on LB agar plates containing Erythromycin (200 µg/ml) supplemented with X-gal (50 µg/ml) and incubated at 37°C over night. Next day white colonies were rescued by restreaking these onto fresh LB- Erythromycin agar plates and incubated at 37°C over night. The next day single colonies of each clone were transferred to liquid LB medium containing Erythromycin (200 µg/ml) and incubated overnight at 37°C with shaking at 250 rpm.

Plasmids were extracted from the liquid cultures using 15 QIAgen Plasmid Purification mini kit (Qiagen, USA) according to the manufacturer's instructions. Five-µl samples of the plasmids were digested with BamHI. The digestion were checked by gelelectrophoresis on a 0.7 % agarose gel (NuSieve, FMC). The appearance of two DNA fragments of about 2.7 kb and 3.7 20 indicated a positive clone. The clone was designated MB293 and the plasmid pMB293.

2)

PMB293 was then EcoRI digested and the 3.8 kb fragment was 25 gelpurified on a 0.7 % agarose gel and using QiaQuick gelpurification kit from Qiagen according to the manufacturer. This fragment which constitutes the pE194 with inserted multiple cloning sites (originating from the PCR primers and from MCS of pUC19) was circulised by ligation and used to transform B.subtilis 30 DN1885, transformed cells were selected on LBPG plates containing 5 µg/ml Erythromycin and incubation was done at 33°C overnight. Clones from this transformation were restreaked and cultivated overnight in TY with 5 µg/ml Erythromycin. From overnight culture broth cells were isolated and plasmids were purified 35 using Qiaquick plasmid kit. Restriction enzyme digestion of the plasmids verified the correctness of the clones. One such clone and plasmid were termed MB333 and pMB333.

3)

The pMB333 was then used as the temperature sensitive plasmid for the construction of a plasmid containing the the Savizyn gene without any promoter. This construct was made as follows.

5

PCR using the primers #21547 and #21548 and the pSX222 as the template was performed according to these conditions:

1  $\mu$ l of pSX222 Qiaquick plasmid prep was used for 50  $\mu$ l  
10 PCR amplification in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM  
KCl, 1.5 mM MgCl<sub>2</sub>, 0.01 % (w/v) gelatin) containing 200  $\mu$ M of  
each dNTP, 2.5 units of AmpliTaq polymerase (Perkin-Elmer,  
Cetus, USA) and 100 pmol of each primer:  
#21547  
15 5'-GCAGCTGCAGGATCCGAATTCGCGGCCGCGCAATCGGTACCATGG-3'  
#21548  
5'-GCAGCGAGCTCAAGCTTCCCGGGCCCAGCCGGTCGACCGCGTTGCCGTTCTGCG-3'

The PCR reactions was performed using a DNA thermal  
20 cycler (Landgraf, Germany). One incubation at 94°C for 1 min  
followed by forty cycles of PCR performed using a cycle profile  
of denaturation at 94°C for 1 min, annealing at 55°C for 1 min,  
and extension at 72 °C for 2 min. Five- $\mu$ l aliquots of the amplification  
25 product was analyzed by electrophoresis in 0.7 % agarose gels (NuSieve, FMC) .

#### Subcloning of PCR fragment

Fortyfive- $\mu$ l aliquots of the PCR products generated as  
described above were purified using QIAquick PCR purification  
30 kit (Qiagen, USA) according to the manufacturer's instructions.  
The purified DNA was eluted in 50  $\mu$ l of 10mM Tris-HCl, pH 8.5.  
Fifty - $\mu$ l of the purified PCR fragment was digested with BamHI  
and ApaI, electrophoresed in 0.8 % low gelling temperature  
agarose (SeaPlaque GTG, FMC) gels, the relevant fragment was  
35 excised from the gels, and purified using QIAquick Gel  
extraction Kit (Qiagen, USA) according to the manufacturer's  
instructions. Purified fragment was ligated to gelpurified BamHI  
and ApaI digested pMB333. The ligation was used to transform

B. subtilis DN1885. Transformed cells were plated on LBPG agar plates containing 5 µg/ml Erythromycin and incubated overnight at 33°C. Clones were restreaked on similar plates and plasmids were isolated from overnight cultures of clones incubated at 5 33°C in TY with 5 µg/ml Erytrhomycin plasmids were isolated using Qiaquick plasmid kit as described by manufacturer. Plasmids were checked by enzymatic digestion using ApaI and BamHI. One correct clone was termed MB339 and the plasmid pMB339.

10

4)

We wanted to test an easy way of monitoring that hybrid plasmids had formed. Thus we wanted to monitor the forced recombination between the two homologous genes on the two seperate plasmids. 15 This was done by constructing a pE194 derivative having a promoter-lacking Savizyn transcrioptionally fused to an ORF encoding GFP. In this way there would be no expression of GFP (due to lack of promoter upstream of the Savizyn and thus the GFP gene) unless the two homologous genes had recombined and thus getting 20 the GFP under control of the promoter otherwise present on pSX120. The plasmid was constructed by PCR amplifying the GFP gene and introducing it just downstream of the savizyn gene on the pMB339.

The GFP encoding DNA used herein is a derivative of 25 the original wildtype gene, the derivative of GFP was cloned from the DNA construction of the mutant F64L-S65T-GFP which was constructed as described in international patent application WO 97/11094.

The construction of the template plasmid was de- 30 scribed in WO 97/11094. The DNA fragment encoding the F64L-S65T-GFP was PCR amplifed from the E.coli plasmid encoding this gene as follows:

1 µl of pF64L-S65T-GFP Qiaquick plasmid prep was used for 50 µl PCR amplification in PCR buffer (10 mM Tris-HCl, pH 35 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01 % (w/v) gelatin) containing 200 µM of each dNTP, 2.5 units of AmpliTaq polymerase (Perkin-Elmer, Cetus, USA) and 100 pmol of each primer:

#20231

5'-AACTGCAGAAGATGTGGACGCGC-3'

#101381

5'-GCCCAATGCATAAACTGCATCCCTTAAC TTGTTTATTGTATAGTTCATCCATGCCATG-  
3'

5

The PCR reactions was performed using a DNA thermal cycler (Landgraf, Germany). One incubation at 94°C for 1 min followed by forty cycles of PCR performed using a cycle profile of denaturation at 94°C for 1 min, annealing at 55°C for 1 min, 10 and extension at 72 °C for 2 min. Five- $\mu$ l aliquots of the amplification product was analyzed by electrophoresis in 0.7 % agarose gels (NuSieve, FMC) .

Subcloning of PCR fragment

15 Forty-five- $\mu$ l aliquots of the PCR products generated as described above were purified using QIAquick PCR purification kit (Qiagen, USA) according to the manufacturer's instructions. The purified DNA was eluted in 50  $\mu$ l of 10mM Tris-HCl, pH 8.5. The PCR fragment and pMB339 were digested with Clal and NsII, 20 gelpurified as above and ligated together. The ligation mixture was used to transform DN1885, transformed cells were plated on LBPG 5  $\mu$ g/ml Erythromycin. After 18 hours of incubation green flourescent cells were visible as single colonies. One of these were analysed and saved as MB406 and the corresponding plasmid as 25 pMB406. Since the GFP encoding ORF did not have an upstream promoter driving the transcription a possible explanation for the GFP expression could be a read through from a promoter further upstream of the Savizyn-GFP transcriptional fusion, namely the promoter of the repF gene.

30

5)

In order two abolish any read-through into the ORF of GFP, a transcriptional terminator was inserted between the repF gene and the Savizyn DNA. The terminator was amplified from the pMUTIN-4-MCS 35 plasmid using the following primers and conditions:

1  $\mu$ l of pMUTIN-4-MCS Qiaquick plasmid prep was used for 50  $\mu$ l PCR amplification in PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.01 % (w/v) gelatin) containing 200  $\mu$ M of

each dNTP, 2.5 units of AmpliTaq polymerase (Perkin-Elmer, Cetus, USA) and 100 pmol of each primer:

#102588

5'-CGGACGGTACCGGTAATGACTCTAGCTTGAGGC-3'

5 #102589

5'-GGGTACAGATCTCCGGCGCAAAAGGCCATCCGTCAGG-3'

The PCR reactions was performed using a DNA thermal cycler (Landgraf, Germany). One incubation at 94°C for 60 sec followed by forty cycles of PCR performed using a cycle profile of denaturation at 94°C for 30 sec, annealing at 55°C for 30 min, and extension at 72 °C for 1 min. Five- $\mu$ l aliquots of the amplification product was analyzed by electrophoresis in 0.7 % agarose gels (NuSieve, FMC) .

The amplified DNA was purified as above and cloned as a KpnI-  
15 BglIII fragment in pMB406 also BglII and KpnI digested (same procedure for gelpurifying etc. as decribed above). The ligation mixture was used to transform DN1885, transformed cells were plated on LBPG 5  $\mu$ g/ml Erythromycin. After 18 hours of incubation cells were visible as single colonies. One of these were ananlysed and  
20 saved as MB430 and correspondingly the plasmid pMB430. This clone had absolutely no expression of GFP indicating that the cloned terminator in fact worked as anticipated.

The resulting sequence of the entire plasmid pMB430 is listed as  
SEQ NO 1.

25

Establishing the MB432 strain

PL1801 *Bacillus subtilis* was transformed with the two plasmids pMB430 and pSX120 using approximately 1  $\mu$ g of each plasmid per 100  $\mu$ l of competent PL1801. The transformed cells were plated onto  
30 LBPG-5  $\mu$ g/ml Erythromycin, incubated overnight at 33°C. Next day colonies were restreaked onto LBPG-5  $\mu$ g/ml Erythromycin + 6  $\mu$ g/ml Chloramphenicol and incubated overnight at 33°C. Clones were also grown i liquid culture TY- 5  $\mu$ g/ml Erythromycin + 6  $\mu$ g/ml Chloramphenicol overnight at 33°C, plasmids were purified from culture  
35 broth and the existence of both pMB430 and pSX120 in the PL1801 strain was verified. This clone was saved and termed MB432.

## EXAMPLE 1

**Sequence shuffling of two homologous sequences both coding for the subtilisin gene of Savinase**

5        The Savinase wt gene (comprised in pSX120) coding for a protease of the subtilisin type was chosen as the target for shuffling of highly homologous genes. A site-directed variant of the Savinase gene (hereafter called Savizyn, comprised in pSX222, where a large number of silent mutations giving new re-10 striction sites are introduced, was used as the shuffling partner to Savinase. The two genes were cloned on two separate compatible plasmids according to the invention. A transcriptionally active Savinase gene was cloned on a plasmid based on the pUB110 origin coding for resistance to chloramphenicol (pSX120, fig. 15 4). A promoter less Savizyn gene was cloned on a derivative of the temperature sensitive plasmid pE194 which also codes for resistance to erythromycin (pMB430). Two transcription terminators were cloned upstream of the Savizyn gene to prevent any read-through.

20       In pMB430 the green fluorescence protein (GFP) was transcriptionally fused to the Savizyn gene in order to be able to visually follow the in-cross and out-cross events between the two plasmids.

25       *B. subtilis* strain MB432, pL1801 transformed with the two plasmids pSX120 and pMB430, were taken through two cycles of temperature shifts to induce two consecutive double-cross over events between the homologous genes. Homologous recombination and plasmid hybrid formation was selected on LBPG plates at the 30 restrictive temperature (47-50°C) in the presence of erythromycin. Under these restrictive conditions, the temperature sensitive plasmid pMB430 is forced into the stable pSX120 plasmid by homologous recombination in the region of Savinase/Savizyn. These colonies where recombination has occurred also becomes fluorescent as a result of the recombination between the two 35 plasmids and concomitant activation of the Savizyn-GFP operon (fig. 2). Out-cross events were screened by growing the clones in LB-broth at the permissive temperature of 30°C for two days and plating the cells on LBPG plate with chloramphenicol and

erythromycin. The non-fluorescent colonies were restreaked on plates and taken through a second round of temperature cycle as described above.

After the second round of shuffling, plasmid preparations 5 were made from 20 different clones. Competent PL1801 was transformed with each of the 20 plasmid preps and plated on LBPG supplemented with erythromycin. The plates were replicated onto LBPG plates with chloramphenicol and erythromycin to identify unwanted clones where both pSX120 and pMB430 had been transformed. Clones where only the pMB430 plasmid had been transformed were selected for sequencing of the Savizyn region to identify if shuffling had taken place. It is evident from figure 6 that recombination between the two genes Savinase and Savizyn had taken place. All of the 12 sequenced clones show some degree 10 of recombination and 4 of these can positively be identified as being result of two consecutive double recombination events. It is important to note that exchange of Savinase and Savizyn sequence can only be observed if the shuffled fragment overlaps a restriction site that is unique for Savizyn. The four clones 15 with a positive evidence for two double recombinations is therefore the minimum number since the recombination could have occurred between two neighbouring restriction sites.

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5313 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: Plasmid pMB430

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GATCAGTACA AGAAAGATAAC TGTATTCAT AAACAGGAAC TGCAAGAAGT TAAGGATGAG	60
TTACAGAAGG CAAATAAGCA GTTACAGAGT GGAATAGAGC ATATGAGGTC TACGAAACCC	120
TTTGATTATG AAAATGAGCG TACAGGTTTG TTCTCTGGAC GTGAAGAGAC TGGTAGAAAG	180
ATATTAAC TGCTGATGAATT TGAACGCCCTG CAAGAAACAA TCTCTTCTGC AGAACGGATT	240
GTTGATGATT ACGAAAATAT TAAGAGCACA GACTATTACA CAGAAAATCA AGAATTAAAA	300
AAACGTAGAG AGAGTTGAA AGAAGTAGTG AATACATGGA AAGAGGGGT A TCACGAAAAA	360
AGTAAAGAGG TTAATAAATT AAAGCGAGAG AATGATAGTT TGAATGAGCA GTTGAATGTA	420
TCAGAGAAAT TTCAAGCTAG TACAGTGACT TTATATCGTG CTGCGAGGGC GAATTCCCT	480
GGGTTTGAGA AAGGGTTTAA TAGGCTAAA GAGAAATTCT TTAATGATTC CAAATTGAG	540
CGTGTGGGAC AGTTTATGGA TGTTGTACAG GATAATGTCC AGAAGGTGCA TAGAAAGCGT	600
GAGAAACAGC GTACAGACGA TTTAGAGATG TAGAGGTACT TTTATGCCGA GAAAACTTT	660

TGCGTGTGAC AGTCCTTAAA ATATACTTAG AGCGTAAGCG AAAGTAGTAG CGACAGCTAT 720  
TAACTTTCGG TTGCAAAGCT CTAGGATTT TAATGGACGC AGCGCATCAC ACGCAAAAG 780  
GAAATTGGAA TAAATGCGAA ATTTGAGATG TTAATTAAAG ACCTTTTGA GGTCTTTTT 840  
TCTTAGATTT TTGGGGTTAT TTAGGGGAGA AACATAGGG GGGTACTACG ACCTCCCCC 900  
TAGGTGTCCA TTGTCCATTG TCCAAACAAA TAAATAAATA TTGGGTTTTT AATGTTAAA 960  
GGTTGTTTTT TATGTTAAAG TGAAAAAAAC AGATGTTGGG AGGTACAGTG ATGGTTGTAG 1020  
ATAGAAAAGA AGAGAAAAAA GTGCTGTTA CTTTAAGACT TACAACAGAA GAAAATGAGA 1080  
TATTAAATAG AATCAAAGAA AAATATAATA TTAGCAAATC AGATGCAACC GGTATTCTAA 1140  
TAAAAAAATA TGCAAAGGAG GAATACGGTG CATTAAAC AAAAAAAGAT AGACAGCACT 1200  
GGCATGCTGC CTATCTATGA CTAAATTTG TTAAGTGTAT TAGCACCCTT ATTATATCAT 1260  
GAGCGAAAAT GTAATAAAAG AACTGAAAAA CAAGAAAAT TCAAGAGGAC GTAATTGGAC 1320  
ATTIGTTTA TATCCAGAAT CAGCAAAAGC CGACTGGTTA GAGTATTAA AAGAGTTACA 1380  
CATTCAATTG GTAGTGTCTC CATTACATGA TAGGGATACT GATACAGAAAG GTAGGATGAA 1440  
AAAAGAGCAT TATCATATTC TAGTGATGTA TGAGGGTAAT AAATCTTATG AACAGATAAA 1500  
ATAAATTAAC AGAAGAATTG AATGCCACTA TTCCGCAGAT TGCAGGAAGT GTGAAAGGTC 1560  
TTGTGAGATA TATGCTTCAC ATGGACGATC CTAATAAATT TAAATATCAA AAAGAAGATA 1620  
TGATAGTTA TGGCGGTGTA GATGTTGATG AATTATTAAA GAAAACAACA ACAGATAGAT 1680  
ATAAATTAAT TAAAGAAATG ATTGAGTTA TTGATGAACA AGGAATCGTA GAATTAAAGA 1740  
GTTTAATGGA TTATGCAATG AAGTTAAAT TTGATGATTG GTTCCCGCTT TTATGTGATA 1800

ACTCGCGTA TGTTATTCAA GAATATATAA AATCAAATCG GTATAAATCT GACCGATAGA	1860
TTTGAATT AGGTGTCACA AGACACTCTT TTTCGCACC AGCGAAAACT GGTTTAAGCC	1920
GACTGCGCAA AAGACGAATT CGAGCTCGGT ACCGGTAATG ACTCTCTAGC TTGAGGCATC	1980
AAATAAAACG AAAGGCTCAG TCGAAAGACT GGGCCTTCG TTTTATCTGT TGTTGTCGG	2040
TGAACGCTCT CCTGAGTAGG ACAAAATCCGC CGCTCTAGCT AAGCAGAAGG CCATCCTGAC	2100
GGATGGCCTT TTTGCGCCGC GGAGATCTAA ATATTCGTGG TGGCGCAAGC TTTGTACCAG	2160
GGGAACCGTC GACTCAAGAT GGGAAATGGGC ATGGCACCGCA TGTGGCCGGG ACGATCGCTG	2220
CTTTAAACAA TTCGATTGGC GTTCTTGGCG TAGCGCCGAG CGCTGAGCTA TACGCTGTTA	2280
AAGTCCTAGG GGCGAGCGGT TCAGGTTCGG TCAGCTCGAT TGCCCAAGGA TTGGAATGGG	2340
CAGGGAACAA TGGCATGCAC GTTGCTAATT TGAGTTAGG AAGCCCTTCG CCAAGTGCCA	2400
CACTCGAGCA AGCTGTTAAT AGCGCGACTT CTAGAGGCCT TCTTGTGTA CGGGCATCTG	2460
GGAATTTCAGG TGCAGGCTCA ATCAGCTATC CGGCGCGCTA TGCGAACGCA ATGGCAGTCG	2520
GAGCTACTGA TCAAAACAAC AACCGCGCTA GCTTTCACTA GTATGGCGCA GGCCTTGACA	2580
TTGTCGCACC CGGGGTAAAC GTGCAGAGCA CATACCCAGG TTCAACATAT GCCAGCTTAA	2640
ACGGTACATC GATGGCTACT CCTCATGTT CAGGTGCGGC CGCCCTTGT AAACAAAAGA	2700
ACCCATCTTG GTCTAATGTA CAAATTGAA ATCATCTAAA GAATACGGCA ACTAGTTAG	2760
GAAGCACGAA CTTGTATGGA AGCGGACTTG TAAACGCAGA AGCGGCAACG CGTTAAAAT	2820
GAGGAGGGAA GCTTTATGAG TAAAGGAGAA GAACTTTCA CTGGAGTTGT CCCAATTCTT	2880
GTTGAATTAG ATGGCGATGT TAATGGCAA AAATTCTCTG TTAGTGGAGA GGGTGAAGGT	2940

GATGCAACAT ACGGAAAACT TACCCTTAAA TTTATTTGCA CTACTGGGAA GCTACCTGTT	3000
CCATGGCCAA CGCTTGTACAC TACTCTCTCT TATGGTGTTC AATGCTTTTC TAGATACCCA	3060
GATCATATGA AACAGCATGA CTTTTCAAG AGTGCATGC CCGAAGGTTA TGTACAGGAA	3120
AGAACTATAT TTTACAAAGA TGACGGGAAC TACAAGACAC GTGCTGAAGT CAAGTTGAA	3180
GGTGATACCC TTGTTAATAG AATCGAGTTA AAAGGTATTG ATTTAAAGA AGATGGAAC	3240
ATTCTGGAC ACAAAATGGA ATACAATTAT AACTCACATA ATGTATACAT CATGGCAGAC	3300
AAACCAAAGA ATGGCATCAA AGTTAACTTC AAAATTAGAC ACAACATTAA AGATGGAAGC	3360
GTTCAATTAG CAGACCATTAA TCAACAAAAT ACTCCAATTG GCGATGGCCC TGTCCTTTA	3420
CCAGACAACC ATTACCTGTC CACGCAATCT GCCCTTCCA AAGATCCCAA CGAAAAGAGA	3480
GATCACATGA TCCTTCTTGA GTTGTAACA GCTGCTGGGA TTACACATGG CATGGATGAA	3540
CTATACAAAT AAAACAAGTT AAGGGATGCA GTTTATGCAT CCCTTAACCTT ACTTATTAAA	3600
TAATTTATAG CTATTGAAAA GAGATAAGAA TTGTTCAAAG CTAATATTGT TTAAATCGTC	3660
AATTCCCTGCA TGTTTAAGG AATTGTTAAA TTGATTTTT GTAAATATTT TCTTGTATTTC	3720
TTTGTAAACC CATTTCATAA CGAAATAATT ATACTTTGT TTATCTTGT GTGATATTCT	3780
TGATTTTTCTACTTAATC TGATAAGTGA GCTATTCACT TTAGGTTAG GATGAAAATA	3840
TTCTCTTGGAA ACCATACTTA ATATAGAAAT ATCAACTTCT GCCATTAAGA GTAATGCCAA	3900
TGAGCGTTTT GTATTTAATA ATCTTTAGC AAACCCGTAT TCCACGATTA AATAATCTC	3960
ATTAGCTATA CTATCAAAAAA CAATTTGCG TATTATATCC GTACTTATGT TATAAGGTAT	4020
ATTACCATAT ATTTTATAGG ATTGGTTTT AGGAAATTAA AACTGCAATA TATCCTGTT	4080

TAAAACCTGG AAATTATCGT GATCAACAAG TTTATTTCT GTAGTTTGC ATAATTATG	4140
GTCTATTC AATGGCAGTTA CGAAATTACA CCTCTTACT AATTCAAGGG TAAAATGGCC	4200
TTTCCTGAG CCGATTTCAA AGATATTATC ATGTCATTT AATCTTATAT TTGTCATTAT	4260
TTTATCTATA TTATGTTTG AAGTAATAAA GTTTGACTG TGTTTATAT TTTCTCGTT	4320
CATTATAACC CTCTTAATT TGGTTATATG AATTTGCTT ATTAACGATT CATTATAACC	4380
ACTTATTTTG TGTTGGTTG ATAATGAAC GTGCTGATTA CAAAAACTA CAAAATGCC	4440
ATATTTTTC CTCCCTATAA AATTAGTATA ATTATAGCAC GAGCTCTGAT AAATATGAAC	4500
ATGATGAGTG ATCGTTAAAT TTAACTGCA ATCGGATGCG ATTATTGAAT AAAAGATATG	4560
AGAGATTTAT CTAATTCTT TTTCTTGT AAAAAAGAAA GTTCTAAAG GTTTTATAGT	4620
TTGGTCGTA GAGCACACGG TTTAACGACT TAATTACGAA GTAAATAAGT CTAGTGTGTT	4680
AGACTTTATG AAATCTATAT ACGTTATAT ATATTATTA TCCGGAGGTG TAGCATGTCT	4740
CATTCAATT TGAGGGTTGC CAGAGTTAAA GGATCAAGTA ATACAAACGG GATACAAAGA	4800
CATAATCAAA GAGAGAATAA AACTATAAT AATAAGACA TAAATCATGA GGAAACATAT	4860
AAAAATTATG ATTGATTAA CGCACAAAAT ATAAAGTATA AAGATAAAAT TGATGAAACG	4920
ATTGATGAGA ATTATTCAGG GAAACGTAAA ATTGGTCAG ATGCAATTG ACATGTGGAC	4980
GGACTGGTTA CAAGTGATAA AGATTCCTT GATGATTAA GCGGAGAAGA AATAGAACGA	5040
TTTTTAAAG ATAGCTTGGG GTTCTAGAA AATGAATACG GTAGGAAAA TATGCTGTAT	5100
GCGACTGTCC ATCTGGATGA AAGAGTCCC CATATGCACT TTGGTTTGT CCCTTAACA	5160
GAGGACGGGA GATTGTCTGC AAAAGAACAG TTAGGCAACA AGAAAGACTT TACTCAATTA	5220

WO 98/28416

PCT/DK97/00567

37

CAAGATAGAT TTAATGAGTA TGTGAATGAG AAAGGTTATG AACTGAAAG AGGCACGTCC 5280

AAAGAGGTTA CAGAACGAGA ACATAAAGCG ATG 5313

Claims

1. A process for *in vivo* recombination of homologous DNA sequences, comprising the following steps

5 a) incubating a cell containing at least two DNA structures, each comprising a DNA sequence and an origin of replication, under conditions which favour integration of one of the DNA-structures into one of the others, thereby forming a hybrid DNA structure;

10 b) incubating the cell under conditions which favour crossing out from said hybrid DNA structure, thereby forming novel DNA structures, each comprising a recombined DNA sequence and an origin of replication;

c) repeating steps a)-b) at least once.

15

2. A process for *in vivo* recombination of homologous DNA sequences, comprising the steps

a) incubating a cell containing at least two DNA structures, each comprising a DNA sequence and an origin of replication, under conditions where i) at least one origin of replication is non-functional and ii) the cell is under a selective pressure which only allows growth of the cell if it contains the DNA structure which is having the non-functional origin (i.e. favoring an integration of one of the DNA-structures into one of the others, thereby forming a hybrid DNA structure);

20 b) changing of said conditions to conditions, where fewer origins of replication are non-functional than in step a) (i.e. favoring a crossing out from said hybrid DNA structure);

25 c) repeating steps a)-b) at least once.

30

35 3. The process according to claim 1-2, characterized in that at least one DNA structure is a chromosome, preferred a bacterial chromosome, more preferred a *Bacillus* chromosome, especially a *Bacillus subtilis* chromosome.

4. The process according to claim 1-3, characterized in that at least one DNA structure is a vector, preferred a bacterial vector, more preferred a vector able to replicate in a *Bacillus* cell, especially a *Bacillus subtilis* cell.

5

5. The process according to claim 4, wherein at least two of the DNA constructs are two different vectors.

6. A process for in vivo recombination of homologous DNA sequences, comprising the steps

- a) inserting the homologous DNA sequences in at least two different vectors, said vectors contain different origins of replication;
- b) transferring the vectors into a cell;
- 15 c) incubating the cell under conditions where i) at least one origin of replication is non-functional and ii) the cell is under a selective pressure which only allows growth of the cell if it contains the DNA structure which is having the non-functional origin (i.e. favoring an integration of one of the DNA-structures into one of the others, thereby forming a hybrid DNA structure);
- 20 d) changing of said conditions to conditions, where fewer origins of replication are non-functional than in step c) (i.e. favoring a crossing out from said hybrid DNA structure);
- 25 e) repeating steps c)-d) as many times as wanted, however at least once.

7. The process according to claim 1-6, characterised in that at least one origin of replication is selected in such a way that the replication rate of the vector can be regulated chemically or physically.

8. The process according to claim 7, characterised in that at least one origin of replication is temperature sensitive.

9. The process according to claim 8, characterised in that the replication rate of at least one vector can be regulated by a temperature shift.

5 10. The process according to claim 1-9, characterised in that the vectors contain genes for markers, preferred selectable markers, especially antibiotic markers.

11. The process according to claim 10, characterised in that the 10 vectors contain genes for different markers.

12. The process according to any of the preceding claims, wherein at least one DNA structure do not comprise ori(-) (i.e. only comprise Ori(+)).

15 13. The process according to claim 12, wherein the DNA structure which do not comprise Ori(-) is a DNA structure which origin (i.e. herein Ori(+)) is functional during the hybrid DNA structure formation performed according a process according to any of 20 the preceding claims.

14. The process according to any of the preceding claims, wherein it is formation of the hybrid structure is measured *in vivo*, preferably by having a system of DNA structures, wherein 25 an *in vivo* screenable protein is expressed only after the hybrid structure has been formed.

15. The process according to claim 14, wherein said screenable protein is a protein which is fluorescent under suitable exposure, and in particular said fluorescent protein is Green Fluorescent protein (GFP) or variants thereof.

16. The process according to any of the preceding claims, wherein crossing out from the hybrid DNA structure is a forced 35 crossing out event.

17. The process according to claim 16, wherein the forced crossing out from the hybrid structure is performed by:

- i) constructing a system wherein the DNA hybrid structure do not express a selectable marker, and at least one of the at least two DNA structures according to any of the preceding claims is able to express said selectable marker;
- 5 ii) incubating the cell under selective conditions, thereby forcing a cross out of the DNA hybrid structure, since the cell is only able to express the marker gene after the crossing out of the DNA hybrid structure has taken place.

10 18. The process according to 17, wherein the selectable marker in step i) is expressed on a DNA structure where the origin of replication is non-functional during a forced integration of the 15 DNA structures into a DNA hybrid structure performed as described above, and preferably the selectable marker gene is an antibiotic resistance gene.

19. The process according to any of the preceding claims, 20 wherein the homologous DNA sequences comprise DNA sequences from at least one pre-made DNA library.

20. The process according to any of the preceding claims, wherein the homologous DNA sequences are comprised of two or 25 more different set of homologous DNA sequences ("set A" "set B", "setC" etc.), and wherein each of said set of homologous sequences ("set A" "set B", "setC" etc.) are inserted in a different DNA construct ("DNA construct A", "DNA construct B", "DNA construct C" etc..) according to any of the preceding claims.

30 21. The process according to claim 20, wherein at least one of said sets of homologous DNA sequences is derived from a pre-made DNA library, and at least one of said sets of homologous DNA sequences is a DNA sequence which are homologous to said DNA library (i.e. resulting in a process for in vivo recombination of 35 said DNA library to a DNA sequence which are homologous to said DNA library).

22. The process according to 20, wherein said sets of homologous DNA sequences is derived from the same pre-made DNA library (i.e. resulting in a process for *in vivo* recombination of said DNA library).

5

23. The process according to claim 20, wherein said sets of homologous DNA sequences are derived from at least two different pre-made DNA libraries (i.e. resulting in a process for *in vivo* recombination of at least two different DNA libraries).

10

24. The process according to claim 20, wherein said sets of homologous DNA sequences each only contain one DNA sequence (i.e. recombination of two or more homologous sequences).

15 25. The process according to any of the preceding claims, wherein two sets of homologous DNA sequences ("set A" and "set B") are recombined *in vivo*, comprising the steps

- a) inserting sequence(s) from "set A" in a vector containing a temperature resistant origin of replication and a gene for a first marker;
- b) inserting sequence(s) from "set B" in a vector containing a temperature sensitive origin of replication and a gene for a second marker;
- c) transferring the vectors into a cell;
- 25 d) incubating the cell at a temperature where both origins of replication are functional, optionally under a selection pressure that favours cells with both marker genes;
- e) shifting the temperature to a temperature, where only the temperature resistant origin of replication is functional;
- 30 f) repeating steps d)-e) at least once.

26. The process according to any of the preceding claims, wherein two sets of homologous DNA sequences ("set A" and "set B") are recombined *in vivo*, comprising the steps

- a) inserting sequence(s) from "set A" in a vector containing a temperature resistant origin of replication and a gene for a first marker;

- b) inserting sequence(s) from "set B" in a vector containing a temperature sensitive origin of replication and a gene for a second marker;
- c) transferring the vectors into a cell;
- 5 d) incubating the cell at a temperature where only the temperature resistant origin of replication is functional, optionally under a selection pressure that favours cells with both marker genes;
- e) optionally, shifting the temperature to a temperature, 10 where both origins of replication are functional;
- f) repeating steps d)-e) at least once.

27. The process according to claim 1-26, characterised in that the cell containing the DNA structures is defective or transitarily inactivated in the enzymatic mismatch repair system or has a reduced level of mismatch repair.

28. The process according to claim 1-27, characterised in that the cell is a Gram positive cell, such as a *Staphylococcus* cell, 20 a *Streptococcus* cell, a *Bacillus* cell, or in particular a *Bacillus subtilis* cell.

29. The process according to claim 1-28, characterised in that the cell is a Gram negative cell, preferred an *Escherichia* cell, 25 especially an *E. coli* cell.

30. The process according to claim 1-29, characterised in that the homologous DNA sequences are encoding polypeptides, especially enzymes, and in particular enzymes such as an amylase, 30 lipase, cutinase, cellulase, oxidase, phytase, and a protease.

31. A method for producing one or more recombinant protein(s) having a desired biological activity comprising:  
a) performing a process for in vivo recombination of 35 homologous DNA sequences according to any of the preceding claims, wherein at least one of the DNA structure(s) is an expression vector;

- b) expressing the numerous different recombinant polypeptides encoded by the numerous different recombined DNA sequences from step a); and
- c) screen or select the numerous different recombinant proteins from step b) in a suitable screening or selection system for one or more recombinant polypeptides(s) having a desired activity.

5

32. A set of at least two vectors, where at least two vectors  
10 are comprising at least two individually different origins of replication able to function in the same cell under permissive conditions, and two homologous DNA sequences encoding polypeptides, said DNA sequences having more than 50% homology, preferable more than 70%.

15

33. A vector system, comprising at least  
a) a vector A containing a DNA sequence encoding a polypeptide and an origin of replication  
b) a vector B containing a DNA sequence homologous to the  
20 DNA sequence in vector A, and an origin of replication that differs from the origin of replication in vector A.

34. The vector system according to claim 33, characterised in that the DNA sequences encodes polypeptides with a biological  
25 activity, especially enzymes.

35. A cell containing a vector system according to claim 33-34.

36. The cell according to claim 35, characterised in that is a  
30 Gram positive cell, preferred a *Bacillus* cell, especially a *Bacillus subtilis* cell.

37. The cell according to claim 35, characterised in that the cell is a Gram negative cell, preferred an *Escherichia* cell, especially an *E. coli* cell.  
35

38. A recombined DNA sequence, characterised in that it is a result of a process according to one or more of the claims 1-30,

or it is encoding a recombinant protein having a desired biological activity, wherein said recombinant protein having a desired biological is produced by the method according to claim 31.

5

39. A process for producing a polypeptide, especially an enzyme, comprising i) growing a cell expressing a polypeptide encoded by a recombined DNA sequence according to claim 38, and ii) purifying said expressed polypeptide.

10

40. A polypeptide, especially an enzyme, characterised in that it is produced by a process according to claim 39.

15

41. A polypeptide, especially an enzyme, characterised in that it is encoded by the recombined DNA sequence according to claim 38.

Figure 1:

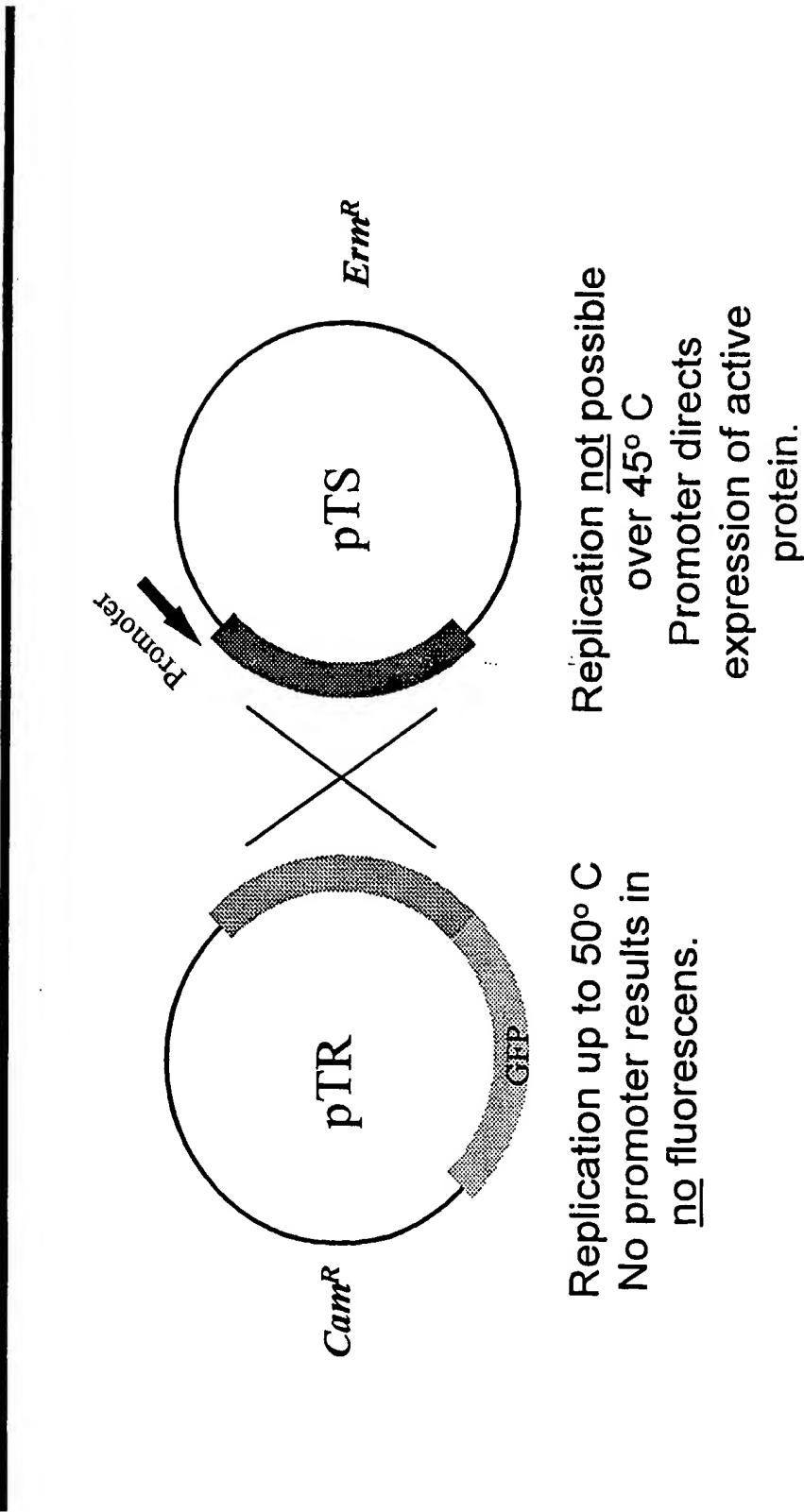


Figure 2:

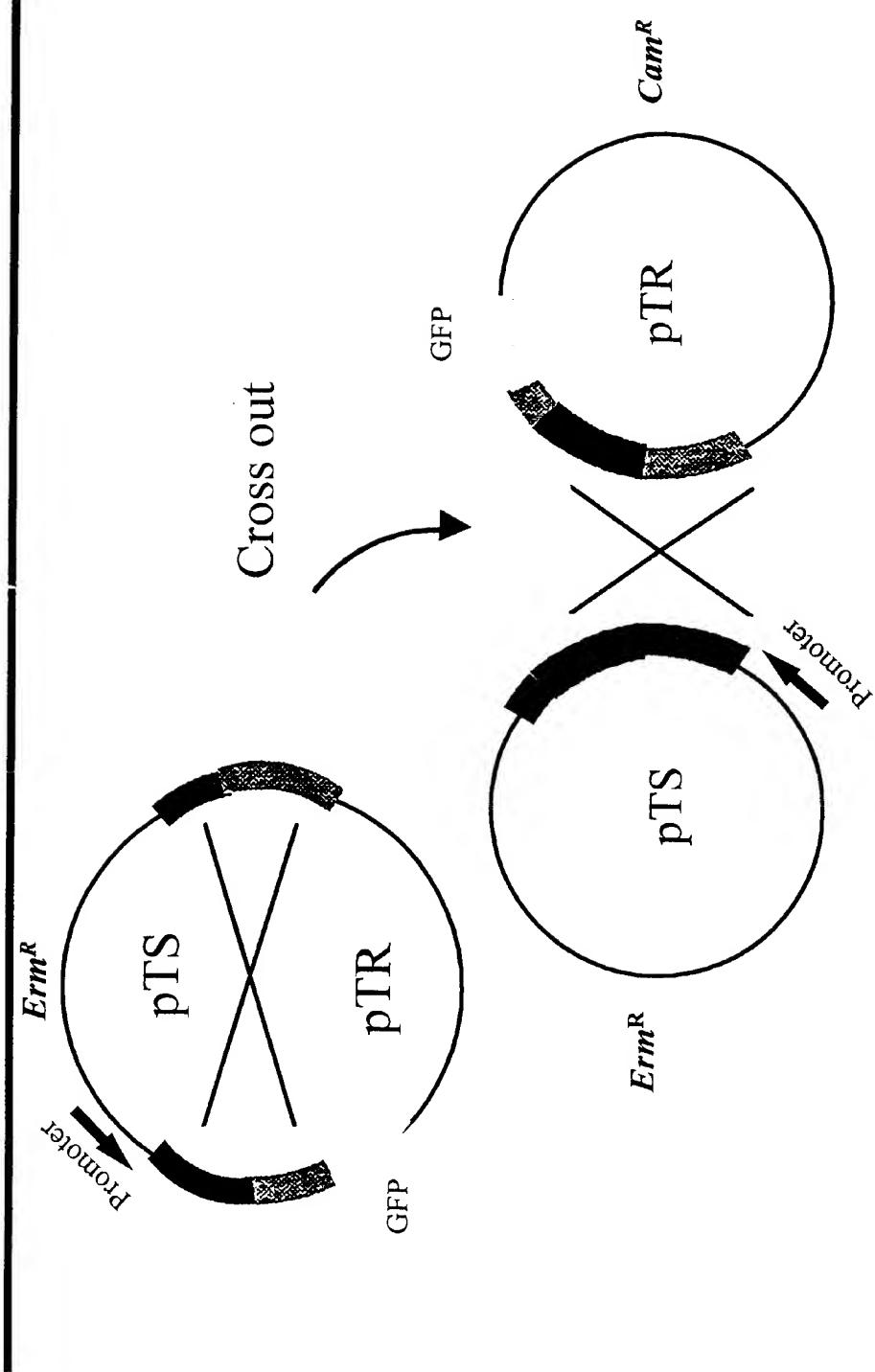


Figure 3:

Shuffled gene.

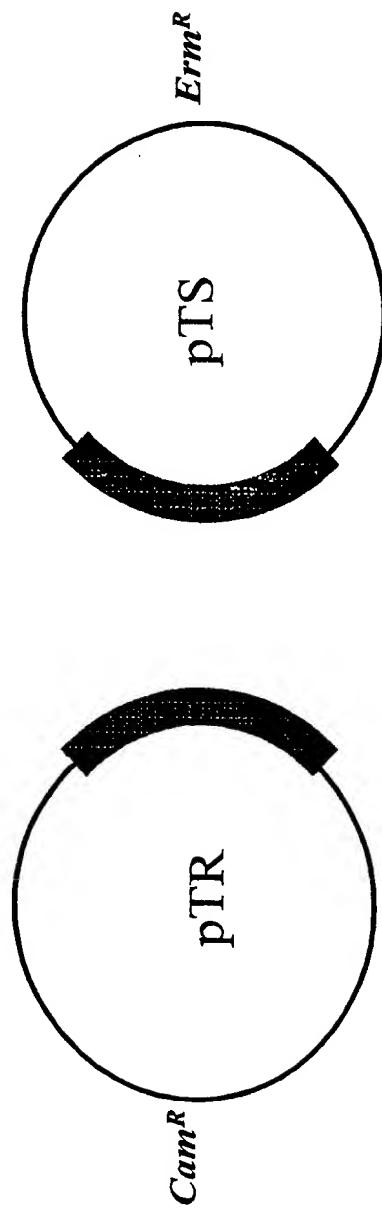


Figure 4:

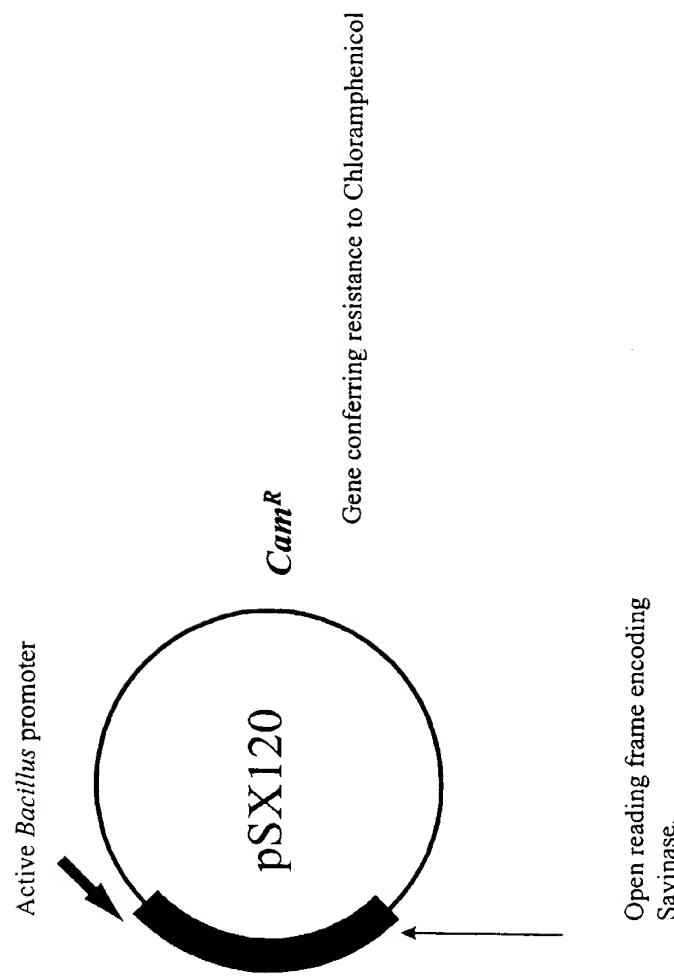


Figure 5:

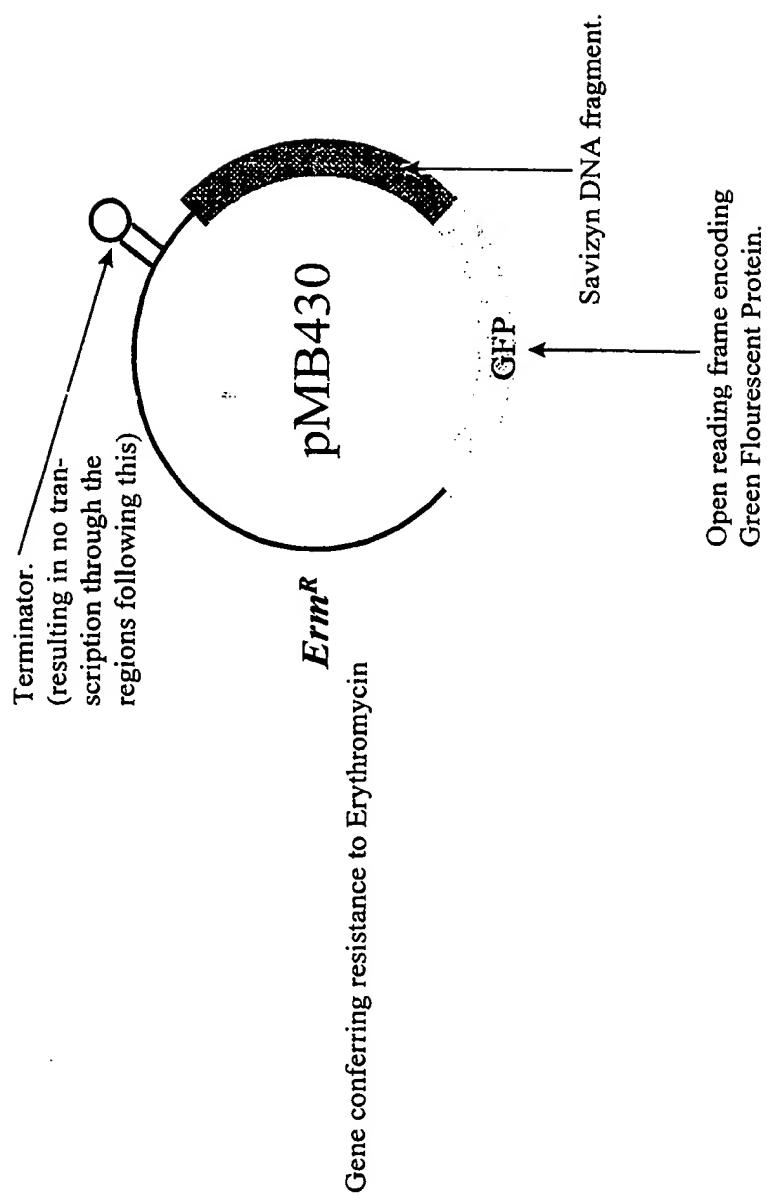
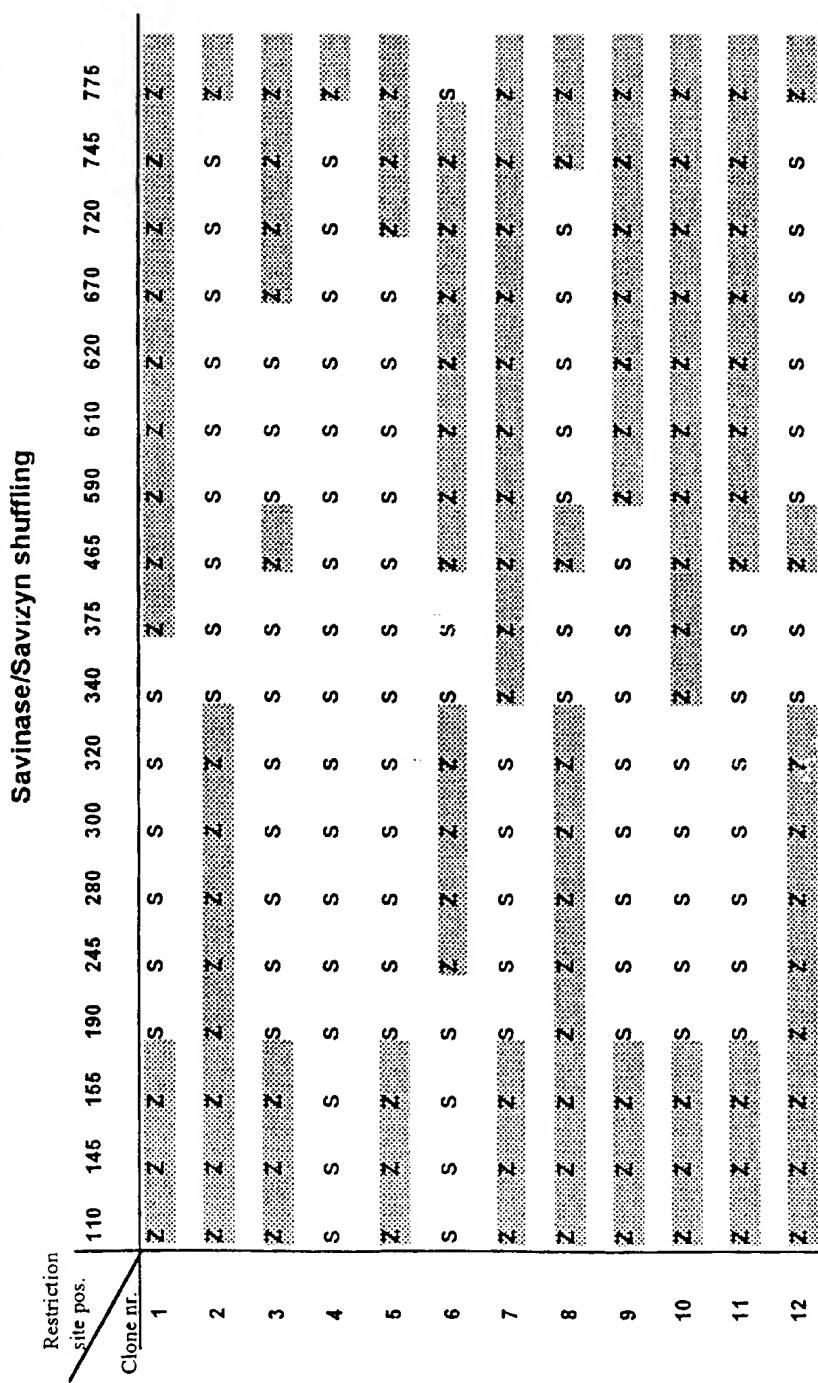


Figure 6:

WO 98/28416

PCT/DK97/00567

6/6



## INTERNATIONAL SEARCH REPORT

**International application No.**

PCT/DK 97/00567

#### A. CLASSIFICATION OF SUBJECT MATTER

IPC6: C12N 15/10

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC6: C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

WPI, EPODOC, BIOSIS, MEDLINE, CA, SCISEARCH

### C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5521077 A (CHAITAN KHOSLA ET AL), 28 May 1996 (28.05.96)	32-37
A		1-31, 38-41
	--	
P,A	WO 9735966 A1 (MAXYGEN, INC.), 2 October 1997 (02.10.97)	1-41
	--	
D,A	WO 9101087 A1 (C.R.C. - COMPAGNIA DI RICERCA CHIMICA S.P.A.), 7 February 1991 (07.02.91), page 15, line 28 - page 23, line 28; page 25, line 14 - page 28, line 23	1-41
	--	

Further documents are listed in the continuation of Box C.

See patent family annex.

- \* Special categories of cited documents:
- "A" document defining the general state of the art which is not considered to be of particular relevance
- "B" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- “T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- “X” document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- “Y” document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- “&” document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 March 1998	06-04-1998

Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +16 8 666 02 86	Authorized officer Patrick Andersson Telephone No. +46 8 782 23 00
--	--

## INTERNATIONAL SEARCH REPORT

International application No.
PCT/DK 97/00567

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Biotechnology, Volume 13, June 1995, Willem P.C. Stemmer, "Searching Sequence Space Using recombination to search more efficiently and thoroughly instead of making bigger combinatorial libraries" page 549 - page 553 --	1-41
A	Science, Volume 263, February 1994, Martin Chalfie et al, "Green Fluorescent Protein as a Marker for Gene Expression" page 802 - page 805 --	15
A	WO 9007576 A1 (SETRATECH), 12 July 1990 (12.07.90), abstract, claims -- -----	27

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

02/03/98

International application No.

PCT/DK 97/00567

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5521077 A	28/05/96	NONE	
WO 9735966 A1	02/10/97	AU 2337797 A AU 2542697 A WO 9735957 A AU 1087397 A WO 9720078 A	17/10/97 17/10/97 02/10/97 19/06/97 05/06/97
WO 9101087 A1	07/02/91	AU 6145190 A	22/02/91
WO 9007576 A1	12/07/90	AT 127519 T AU 4834390 A CA 2006549 A DE 68924174 D,T EP 0449923 A,B SE 0449923 T3 ES 2077058 T FR 2641793 A,B IE 72469 B JP 4503601 T	15/09/95 01/08/90 26/06/90 14/03/96 09/10/91  16/11/95 20/07/90 09/04/97 02/07/92